Quantitative Geography

Alan T. Murray

School of Geographical Sciences
Arizona State University
Tempe, AZ 85287-5302, USA

(Email: atmurray@asu.edu)

Prepared for:
Journal of Regional Science's 50th Anniversary Conference
April 23-24, 2009
Federal Reserve Bank of New York

April 17, 2009

Abstract

This paper provides an overview of quantitative geography, and the methods that have come to define it. This is part of a conference marking the 50th anniversary of the Journal of Regional Science. Six broad categories are used to discuss the range of methods found in quantitative geography: geographic information systems; airborne sensing (global positioning system, photogrammetry and remote sensing); statistics and exploratory spatial data analysis; mathematics and optimization; regional analysis; and, computer science and simulation. Particular emphasis is given to the state of the art in each area, with discussion on major unresolved issues and future research directions.

Introduction

There have been many overview texts on quantitative geography, most somewhat dated at this point. Perhaps the most recognized is that of Wrigley and Bennett (1981), at least in terms of citations in the literature, but others include Taylor (1977) and Cole and King (1968). Of course, there are also the works cast under the headings of "spatial analysis" (Berry and Marble 1968), "statistical analysis" (King 1969, Clark and Hosking 1986) and "locational analysis" (Haggett et al. 1966) that could/should be included in this area as well, certainly touching upon major components of quantitative geography.

Quantitative geography has been defined by Fotheringham et al. (2000) as consisting of "… one or more of the following activities: the analysis of numerical spatial data; the development of spatial theory; and the construction and testing of mathematical models of spatial processes." This is a reasonable definition/characterization, though the creation of spatial information and knowledge may be missing. Further, there is some ambiguity about whether methods developed by other disciplines (mathematics, statistics, engineering, social sciences, etc.) for aspatial contexts would qualify as quantitative geography methods. The fact is that may classic mathematical and statistical methods have been characterized as part of quantitative geography. Thus, one might say that quantitative geography is the collection of methods that are applied, or could/can be applied, by geographers and others to study spatial phenomena, issues and problems. Whatever the case, it is clear that geographers have relied heavily on classic quantitative methods, and have developed rather important extensions in the study of spatial problems and issues. In addition, researchers in other academic disciplines have also contributed to many methods in quantitative geography, making it difficult to attribute many methods to any one discipline.

Given this as a beginning point, the goal of this paper is to provide an overview of quantitative geography, recognizing that such methods have been or could be applied to human and physical geography problems and issues. This spans behavioral patterns, cognition, population forecasting, migration, demography, climate change, sustainability, hydrology, transportation, and many other areas.

One could begin with the so called quantitative revolution, though it was not something unique to geography. It is clear, however, that much of what quantitative geography (and regional science) is can be related back to developments originating in the 1950s, 60s and 70s. Rather than revisit such connections and issues, the interested reader is directed elsewhere for further discussion (e.g., Fotheringham et al. 2000, Haggett 2008). What this overview will do is recognize that a major facet of quantitative geography is development of some sort of model, generally having a significant spatial context or component. Of course, models reflect relationships and knowledge (Rey 2001), and as a result may be overly simplified, inaccurate or incomplete.

There are admittedly many ways to conduct an overview of quantitative geography. It could focus on methods, individuals, methodological nuances of descriptive vs. normative, measures vs. models, etc. What is attempted here is to derive major overarching categories to discuss methods in quantitative geography. There will no doubt be omissions, and most certainly the treatment and discussion is limited. It is inevitable that this treatment may have a human geography and GIScience slant, though attempts are made to reflect all areas of geography. Finally, most of the discussion is intentionally directed to the methods that have been developed and/or extended/enhanced by geographers given the theme of quantitative geography, and necessarily limiting details associated with classic methods from other disciplines generally restated to address geographic/spatial problems.

Primer of Methods in Quantitative Geography

There are a range of techniques and methods used in geographical research and application that fall under the heading/umbrella of quantitative geography. As noted previously, ultimately quantitative geography is the collection of methods that are applied, or could/can be applied, by geographers and others to study spatial phenomena, issues and problems, often over time. This section presents an overview and description of methods in quantitative geography. There is no particular intended ordering, though the discussion naturally builds upon the methods previously discussed. What is true is that many categories rely on other categories in various ways. Thus, many methods could be discussed in multiple categories. Such duplication is avoided due to space limitations.

The following broad categories are used to detail methods in quantitative geography: geographic information systems; airborne sensing (global positioning system, photogrammetry and remote sensing); statistics and exploratory spatial data analysis; mathematics and optimization; regional analysis; and, computer science and simulation. Such a primer review will necessarily be limited, likely overlooking what some may consider important methods, and lacking in providing important theoretical and application details. Included references will hopefully provide a start for those interested in more such details.

Geographic Information System

Arguably the most significant "method" in quantitative geography is a geographic information system (GIS). Formally defined, a GIS is a collection of hardware, software and associated procedures to support spatial data acquisition, management, manipulation, analysis and display (Longley et al. 2005, Church and Murray 2009).

The evolution of GIS is an interesting story, but of significance here is that present day commercial systems are founded on advances in computing, computer programming, mathematics, geodetics, cartography and mapping, as well as other fields of research. A specific example is that GIS has direct ties to the development and use of a database management system (DBMS), a product of computer scientists, but also disciplines and specialty areas concerned with the representation of the earth in a digitial environment (geodetics) and effective communication through maps (cartography and mapping).

In general terms, GIS provides an analysis environment for integrating layers of spatial information in a common coordinate system, where the layers of information are in either raster or vector formats. A raster GIS format reflects a continuously varying surface represented by regularly sized cells that completely cover an entire area. Alternatively, a vector GIS format corresponds to discrete objects (traditionally, points, lines and polygons) in space, representing a non-exhaustive and selective sample of geographic entities.

The major components of GIS (data acquisition/input, management, manipulation, analysis and display) are illustrated in Figure 1. Each component provides a unique and important operational functionality. It is the effective interaction and linkage of these components that enables knowledge to be gained. Discussion of each component is now given.

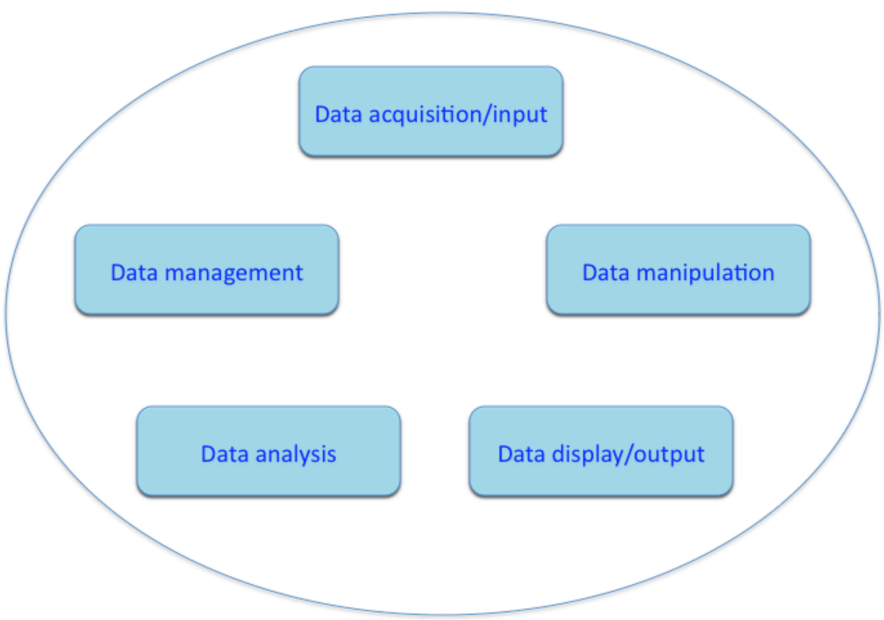

Figure 1. Components of GIS.

Data acquisition/input in GIS concerns capabilities for working with geographic information in a digital environment. This includes the ability to import data in proprietary or industry standard formatted attribute and shape/geography files. Alternatively, acquisition/input support also means that creating digital information is a necessary component of GIS, like the ability to digitize features on a map or scan features/attributes for an area. Related is the capacity to geocode address based information. Finally, input support entails the ability to accommodate external data sources, such as surveying, GPS, areal photographs and remotely sensed imagery, which are discussed below.

The *data management* component of GIS has been and continues to be a rather critical issue as efficient storage and quick access to spatial information and associated attributes is paramount. The success and significance of GIS likely can be attributed to timely response and processing in the management of large quantities of geographic information. This has meant that raster and vector based data be treated uniquely, as this enables certain types of efficiencies to be realized. For example, with raster data the use of a scan order means that only attribute information need be stored for each cell, and that data compression is possible through a process known as run length encoding.

The ability to *manipulate* spatial information in GIS is critical. Basic manipulation is often conversion between coordinate systems using complex affine transformations, but also projection of data from 3-D to 2-D (and vice versa). Another type of manipulation is aggregation, where objects are combined based upon spatial proximity in order to decrease the quantity of data, protect the privacy of individuals/families or better reflect geographic properties. Finally, a third type of manipulation is overlay, where non-coincident layers (the polygon units and associated boundaries are not the same between two or more layers) are combined together to create a new layer containing all uniquely defined spatial units and attributes of the input layers. Thus, a necessary part of overlay is areal interpolation, the process of intelligently deriving attribute values for a portion of a polygon.

Display in GIS has proven to add the wow factor to this method, enabling map based graphics to be easily generated for evaluation and inspection by humans. This is where knowledge is typically derived. Interestingly, this is far more complicated and involved than one may realize, as substantial research continues to be devoted to display oriented endeavors. From the human perception and cognition side, there are issues of appropriate communication in color selection, symbology, etc. Even the most basic choropleth map displays, where polygons are color coded to represent some attribute interval, are involved, with the default natural breaks approach reflecting a class selection mathematical optimization problem. Beyond this, there are issues of uncertainty in spatial data, temporal changes, 3D rendering, and others, that remain current topics of academic research.

The *analysis* component of GIS is either woeful or profuse, depending on your perspective. The view that there is no real analysis capability in GIS other than simple query and map based views of geographic data is no doubt exaggerated, particularly given all that goes into being able to query and generate a map based view, as detailed above. Basic analysis features of GIS that should be highlighted are query (aspatial and spatial), map algebra, buffering, computational geometry functions like creating a voronoi diagram, and point interpolation. Map algebra is rules and operational procedures applied to an attribute layer(s), traditionally raster data, in order to produce a new layer. Such a process was performed by McHarg (1969), where data layers were processed, then overlaid to represent a suitability surface through which a transportation route was to be selected. Buffering is the geometric process of taking an input object(s), point, line or polygon, and extending out some specified distance, resulting in a new polygon based object. Other geometrically based GIS functions include creating voronoi diagrams for a generator point set and identifying trade areas based on travel times, viewsheds, watersheds, etc. A final GIS analysis feature to be mentioned is point interpolation (could note areal interpolation discussed above here too as it is conceptually and technically different), which is the process of intelligently estimating an attribute value at an unsampled location using observed attribute values at other sampled locations.

In summary, there is much to what is considered GIS, building on contributions from many disciplinary fields including geography. While geography cannot claim complete ownership of GIS, it has been a significant contributor to GIS development and application.

Airborne Sensing

The airborne sensing category of methods is a grouping of global position system, photogrammetry and remote sensing, geospatial sensing technologies that can be utilized for data collection and/or creation (Thurston et al. 2003). As with GIS, these technologies reflect the many contributions of many academic disciplines, especially engineering and geodetic sciences. Geography has and continues to make important contributions to these technologies, and they have contributed to the significance of GIS. Each is recognized as important specialty area with unique methods.

A *global position system* (GPS) is a constellation of satellites orbiting the earth, base stations (in the case of differential correction) and individual receivers. Combined, the satellites, base stations and receivers enable position and time on the surface of the earth to be determined. This is accomplished through satellites emitting coded radio signals that can be detected by base stations and receivers, thereby allowing for position and time on the earth’s surface to be established, provided that four or more satellite signals can be acquired from the associated location. The U.S. Department of Defense operates the NAVSTAR GPS, but a similar constellation is GLONASS, operated by the Russian Federation Ministry of Defense. Receivers can therefore be used to identify the geographic coordinates of a location as well as add to this data associated with attribute information, which means that point objects, line objects and area/polygon objects can be created in an automated fashion. GPS derived data is typically of high accuracy, down to a few centimeters, but depends on signal strengths and environmental conditions.

Another areal based data collection approach is *photogrammetry*, where areal photographs are taken from an airborne platform, including balloons, planes and helicopters, and involves issues of data capture, analysis and interpretation. A camera is needed, and subsequent processing must interpret the objects and information contained in the photo as well as taking into account height and vertical angles in deriving coordinate information. It is also necessary to georeference areal photo information on the surface of the earth. It is possible to attain positional accuracy down to a millimeter or so. Given height and vertical angle issues, there is necessarily various types of distortion possible in derived data.

The final airborne approach to be discussed is *remote sensing*. In general terms, remote sensing refers to both hardware and software for semi-automated spatial data collection, where a satellite is relied upon to detect electromagnetic radiation using passive optical and/or microwave sensors. An example is shown in Figure 2. This makes it possible to determine land cover composition, particularly land uses and vegetation types, once the area based measures by sensors are processed and interpreted. In contrast to photogrammetry, remotely sensed imagery generally corresponds to a larger area and has coarser spatial resolution. There are numerous government and commercial remote sensing options, two prominent being Landsat (US) and SPOT (French).

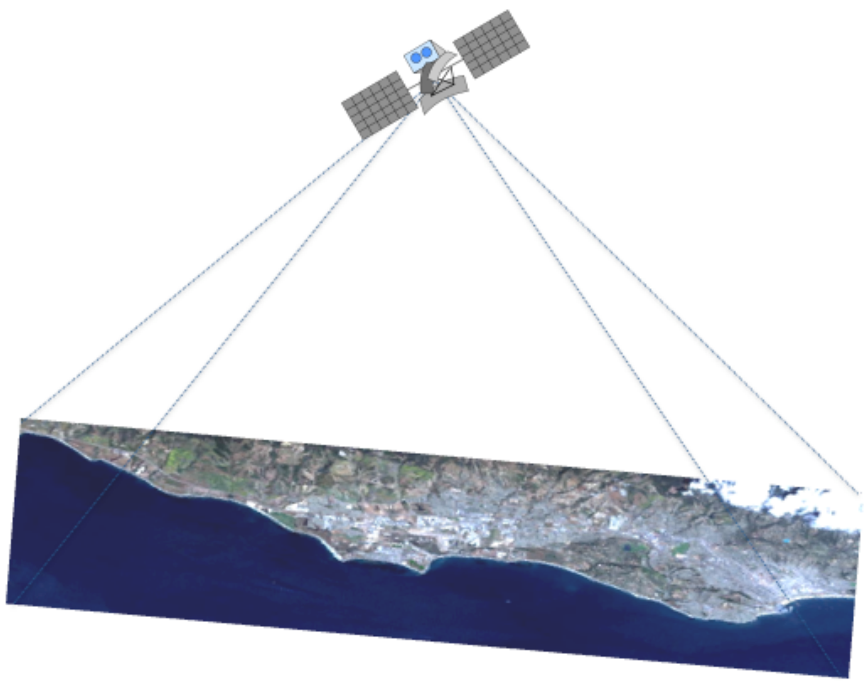

Figure 2. Airborne data collection by satellite.

Mathematics and Optimization

Directly or indirectly, all methods in quantitative geography rely on mathematics of some form. Essential no doubt are the basics of algebra, geometry, calculus and linear algebra, but also more advanced topics, like Fourier analysis, differential equations, Laplace transformations, numerical methods, complex analysis, etc., are integral to the approaches of many human and physical geographers. Further such methods are fundamental in optimization approaches like linear, integer and dynamic programming as well as heuristic methods to solve these problems. In what follows, three prominent areas where mathematics and optimization have been adapted to inherently quantitative geographic methods are reviewed: spatial interaction, spatial optimization and network analysis. Worth noting is that the section on statistics could be included here as well given it mathematical foundations and reliance on optimization techniques, but is left as its own topic because of the many sub-areas that are the focus of geographic research efforts.

There has been considerable work by geographers to develop and apply *spatial interaction* models. Fotheringham et al. (2000) note that spatial interaction originated from social physics and statistical mechanics, and has matured into a specialty area of its own based on theoretical and empirical advances. A rather important advance was the model of Huff (1964) to delineate trade areas, based on the underlying gravity model. What followed as a considerable body of research, amounting to behavioral, cognitive and spatial information processing advances of the basic gravity model (see Fotheringham et al. 2000).

A second area is *spatial optimization*, where an underlying model reflects spatial processes and desires. For example, a linear programming problem can be stated as follows:

$$\textit{Maximize} \quad cx \tag{1}$$
$$\textit{Subject to} \quad Ax \leq b \tag{2}$$
$$x \geq 0 \tag{3}$$

where c is a 1 by n vector of benefits, x is an n by 1 vector of decision variables, A is a m by n matrix of constraint coefficients, and b is a m by 1 vector for right hand side limits. Spatial optimization work has focused on cases where objectives and constraints in such a model, (1) and (2), are inherently geographic, generally with the decision variable vector corresponding to decisions regarding where a good or service should be sited (location modeling), the routing of a corridor, land uses, reserve sites to be selected, and the like (Church and Murray 2009). Murray (1999) structures a harvest scheduling model that is inherently spatial, emphasizing that neighboring or local impacts be limited. Other examples can be found in location analysis and modeling, and are numerous: the location set covering model, the maximal covering location model, dispersion models and the p-median problem, just to name a few. The confounding issues are appropriately and accurately structuring the model of interest with respect to spatial considerations and actually solving the model using either exact and heuristic approaches. Not only this, there remain fundamental issues in the treatment of space and spatial relationships. For example, Murray (2005) highlights that the location set covering model encounters representational issues when applied to the coverage of area based data, necessitating a complete restructuring of the underlying spatial optimization model.

A third area where mathematics and optimization have been substantial is *network analysis*. Broadly conceived, network analysis involves deriving attribute and performance characteristics of an interconnected system of nodes and arcs. Examples of networks are roads, rivers, electrical services, telecommunication systems, etc. Quantitative geography in this area has consisted of the development and use of basic metrics to help characterize spatially oriented elements of a network, like cyclomatic number, beta index, alpha index, gamma index and average length of shortest path (Haggett and Corley 1969). This has evolved to consider shortest path variants in a network (Church and Murray 2009) as well as system connectivity, flow and vulnerability/reliability (Grubseic et al. 2008), generally involving the use of a spatial optimization model of some sort. Thus, network analysis reflects the use of mathematical measures and models to better understand the geographic structure of networks, as well as plan for and protect the efficient continued operation of network based systems, like oil and water pipelines, transportation systems, supply chains, etc.

Statistics and Exploratory Spatial Data Analysis

It is often the case that quantitative geography is equated to the application of statistical techniques to geographic problems. This is likely due to the fact that there are many texts devoted to statistics for geographers, like Clark and Hosking (1986) and Rogerson (2006), but a host of others as well. While this paper highlights that the range of methods is far more encompassing, it remains that statistical measures and methods are an important part of quantitative geography. In the discussion that follows, the range of statistically oriented techniques are touched on that have generally been applied, extended and developed to address geographic issues.

Surveying and sampling have been widely relied on in geographic research. This ranges from qualitatively oriented semi-structured surveys targeting geographic areas and geographic issues to large scale opinion and marketing studies. The appeal and usefulness of sampling is that we can accurately and correctly infer things about a population from a much smaller sample, and it turns out that spatially representative samples are essential in many cases and contexts (Berry and Baker 1968). Thus, spatially representative samples must be a part of an overall study design, and requires a planned approach to achieve sufficient representation.

As with many disciplines, classic descriptive statistical measures (mean, variance, higher ordered moments, correlation, etc.) have been an important part of quantitative geography, as have classic statistical models (regression, analysis of variance, principle components, factor analysis, multidimensional scaling, etc.) and non-parametric approaches (see Wrigley and Bennett 1981). More specific and unique to quantitative geography are the following spatial statistical sub-areas: point pattern analysis, spatial autocorrelation, spatial statistical models, and exploratory spatial data analysis (ESDA). Each will now be discussed.

The area of *point pattern analysis* consists of a number of recognized methods structured to support the analysis of a hypothesis about the spatial distribution points, like whether they are clustered or not, or conform to a particular distribution. Many of the texts noted previously provide detailed discussions on one of the methods that follow (e.g., Taylor 1977, Fotheringham et al. 2000, Rogerson 2006). One method is nearest neighbor, which examines the distance between an observed point and its closest neighboring point. With the distance for all points, a measure of spatial dispersion is constructed as a ratio of observed and expected values for the region. Another approach is the quadrat method, where the study area is divided into a discrete number of cells of equal size. The number of points in each cell is then determined and compared to the hypothesized (or expected) number of points per cell. This is done using a chi-square goodness-of-fit based statistic. A third point pattern approach examines kernel density. This extends the notion of the quadrat method to include counts for neighboring cells within a prespecified radius, and are generally analyzed as a map based surface. A fourth approach is using k functions, which is a so called second order process, in contrast with kernel density, accounting for observed points within a prespecified radius divided by the expected number of points in this area. A final group is clustering methods, and are generally relied upon in spatial data mining. There are both hierarchical and non-hierarchical clustering methods, but optimization based non-hierarchical methods have generally been applied. Clustering approaches identify groups of points that are most similar. A popular approach is the k-means approach, but more spatially correct and refined clustering approaches for spatial point patterns are those detailed in Murray and Estivill-Castro (1998) and Murray and Shyy (2000).

The area of *spatial autocorrelation* recognizes that a variable in areal unit data, in contrast to point based data above, may be correlated with respect to space. In particular, there may be some spatial arrangement or configuration associated with the distribution of that variable. To assess whether this is the case, many methods have been developed, and can be viewed as either global or local (Anselin 1995). A global measure tests for correlation across the entire study region, whereas a local measure focuses on a particular sub-area of the region. One global measure of spatial autocorrelation is the join count, and assumes that the variable is binary (e.g., 1 or 0, where 1 indicates that the attribute/variable exists and 0 that it does not exist in the spatial unit). The statistic then enables the hypothesis of a random pattern to be tested. Continuously measured data is approached along similar lines, a popular approach being Moran's I:

$$I = \frac{z'Wz}{z'z} \tag{4}$$

where z is the n x 1 vector of mean standardized variable values (z' the transpose of z), and W is the so called spatial weights matrix, n x n in dimension. The numerator measures variability from the mean, and the denominator is the variance. Values of I range from 1 (positive spatial autocorrelation) to -1 (negative spatial autocorrelation). Variants of the global measure of spatial autocorrelation take into account alternative ways to measure difference between the attribute value of neighboring spatial units. For example, Geary's c examines difference squared, and the G statistic considers the product. One issue with such measures, however, is that they only indicate that spatial autocorrelation is or may be present, and not where geographically it is significant, not to mention the assumption of spatial stationarity (constant mean and variance across space). To address this, these and other global measures can be and have been localized to consider a spatial unit and its neighbors (Anselin 1995), enabling the identification of so call hot spots (or clusters) of positive and negative spatial autocorrelation.

Not unrelated to spatial autocorrelation is the development of *spatial statistical models* structured to deal with or capable of dealing with spatial dependence (a functional relationship between what occurs at one location and what happens everywhere else) and spatial heterogeneity (a lack of geographic uniformity). Oriented toward dealing with spatial dependence, there is the spatial lag, or mixed regressive, spatial autoregressive model (Anselin 1988):

$$y = X\beta + \rho Wy + \varepsilon \tag{5}$$

where y is the n x 1 vector of observations for the dependent variable, X is the n x k matrix of observations for the k independent variables, β is k x 1 vector of regression coefficients, ρ is the autoregressive coefficient/parameter, W is the n x n spatial weights matrix, and ε is the n x 1 vector of random error terms. Models designed to deal with spatial heterogeneity are the expansion method (Casetti 1972) and geographically weighted regression (GWR) (Brunsdon et al. 1998). The spatial statistical models are effectively multivariate regression approaches the deal with important spatial conditions that violate classic statistical assumptions.

Though mentioned previously, there are also a bevy of interpolation approaches, reflecting spatial statistical models of continuously varying data, in contrast to discrete areal units. The work of Tobler (1979) is particularly noteworthy.

The area of *exploratory spatial data analysis* (ESDA) reflects the need for a process of generating insights on patterns, trends and associations in spatial information, where there is generally a lack of prior assumptions or insights about substantive context or geographic region. ESDA extends exploratory data analysis attributed to Tukey (1977), focusing on aspatial and spatial effects, like spatial distributions, spatial outliers, spatial patterns and spatial regimes of statistical instability (Anselin and Bao 1997). Associated ESDA methods can be basic metrics and statistical measures, but also graphics and map based displays. On the graphs and map based display side, ESDA has been oriented toward the use of dynamic graphics with brushing and linking between displays, like that displayed in Figure 3. This is done through dynamic integration, where information is efficiently and seamlessly moved/passed between graphic and map based displays. The left window in Figure 3 displays a count based graphic summary of origin standardized spatial movement vectors in terms of distance and direction. This is linked to the right window in Figure 3 showing the movement vectors, with the highlighted section of left graph also highlighted in the right window. In terms of the different types of displays possible, Fotheringham et al. (2000) list the following exploratory approaches: stem and leaf plots, box plots, histograms, density surfaces, maps, scatterplot matrix, parallel coordinate plots, radviz and projection pursuit. Anselin and Bao (1997) suggest that ESDA is often directed at understanding spatial distributions (maps and integrated box plots) and spatial association (variogram clouds, variogram boxplots, spatial lag scatter plot for continuous data and spatial lag pies, spatial lag bar charts, Moran scatterplots, mapping significant LISA statistics and the spatial correlogram for areal unit data). A number of ESDA oriented packages have been developed in recent years, including GeoDa (Anselin et al. 2006), STARS (Rey and Janikas 2006) and PySAL (Rey and Anselin 2007).

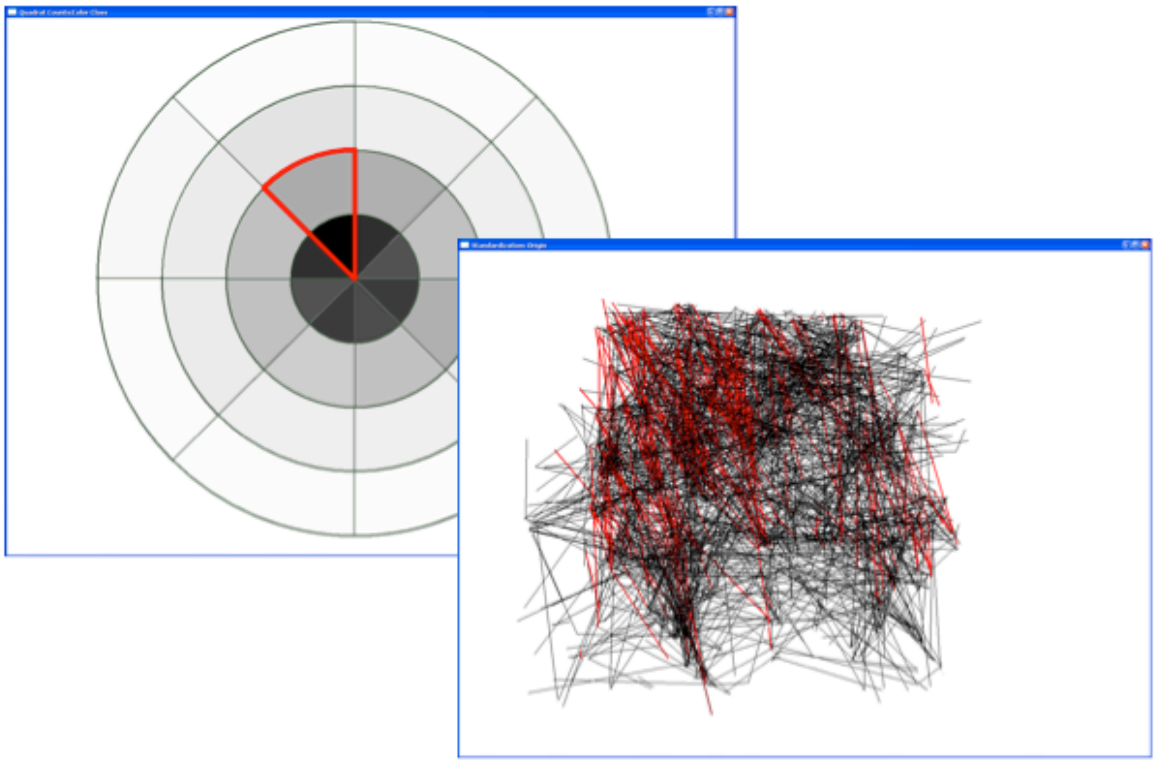

Figure 3. ESDA interactive and linked displays.

Regional Analysis

The category of regional analysis has reflected the need to understand urban and regional economies in a formal, quantitative manner. This is in terms of explaining what is currently taking place, but also why changes occurred and what changes are likely in the future. Many classical economics based methods have been relied upon for regional analysis, including the fundamentals of location theory (land rent models, cost minimization, central place hierarchies, and competition), input-output models, and computable general equilibrium models. These methods and others have been widely applied by geographers, but also extended in various ways.

One method in regional analysis is a simple measure called the location quotient, measuring the relative importance of an industry/sector in a region to a national fraction. Another rather simple approach is shift-share analysis, where a regions employment, as an example, can be divided into three components: national share, regional share and industry mix share. Haynes and Dinc (1997), among others, have discussed extensions and interpretation issues.

A rather prominent technique in regional analysis is the input-output model:

$$X = (I - A)^{-1} Y \tag{6}$$

where X is the n by 1 vector of outputs by industry, I is the n x n identity matrix, A is the n x n matrix of interindustry flow relationships and Y is the n x 1 vector of final demand. It represents an analytical framework for looking at the interdependency of industries in an economy. Much work has looked into various kinds of extensions and interpretive issues, like multi-region and regional effects (Hewings 1985). One could also note issues of spatial and regional decomposition as well as error and sensitivity. For example, the work of Jackson and Murray (2004) focuses on the minimization of information loss in updating interindustry flow data.

One final comment is that we could include the discussion of regional models based on spatial interaction or optimization (e.g., Wilson 1974) in this section, but they have already been included in the previous category of mathematics and optimization.

Computer Science and Simulation

The final category in quantitative geography is computer science and simulation. These two broad disciplinary areas are generally the playground of non-geographers. However, with GIS and geographic models has arisen the need for geographers to make significant contributions in database design, algorithm design and simulation processes. Armstrong (2000) discusses some contributions in this area as well as future potential, but begins by noting that computational science is the use of computing technology to create knowledge. With this in mind, geographic analysis continues to face substantial computational complexity issues, including large data volumes and computationally intensive methods (like those reviewed above). But much potential lies in taking advantage of future computing advances, distributed networks and parallel processing. In what follows, three areas are noted where geographers have made significant contributions: spatial database design, algorithm development and simulating spatial processes.

Mention was made in the discussion of GIS about the importance of spatial data management, and it is precisely the geographic nature of data in GIS that makes it challenging to deal with, both in terms of storage and processing efficiency (*database design*). Further discussion of this issue and contributions from geographers can be found in Longley et al. (2005), and other GIS oriented books and journal articles.

In many, or most, of the above quantitative geography categories, one could have included a discussion on *algorithms*. In general terms, an algorithm is a process for deriving a solution to a model that consists of a finite number of steps. In optimization, it is added that an algorithm also gives an exact solution in the sense of being provably the best, in contrast to a heuristic (Church and Murray 2009). One can think of many algorithms in either sense (exact or heuristic), like natural breaks in choropleth mapping, the shortest path algorithm, creating a TIN (triangulated irregular network) and spatial optimization model heuristics. Church and Murray (2009) note that the interchange approach is a popular heuristic for solving the p-median problem (a spatial optimization model), as an example. It is used in the ArcGIS LocationAllocation package to solve this and related location models. Other algorithms (or rather heuristics) are possible as well, like simulated annealing and tabu search (Murray and Church 1996).

An area of continued importance is *simulating spatial processes*, at both micro and macro levels. Examples include developing simulation approaches to mimic regional growth patterns over time, as well as people, using cellular automata and agent based techniques (Batty and Longley 1994, Clarke et al. 1997, Ward et al. 2000). Another example is the use of neural networks and artificial intelligence (see Fischer 2006). Interestingly, there may or may not be an underlying mathematical model and there may not be any optimizing process, though relationships and changes can be quantified mathematically. It is precisely the processes of change and spatio-temporal relationships that turn out to be effective in predicting change, thereby dictating what will happen in the near and longer term. Such processes and relationships are therefore structured in a computer program to simulate change and behavior.

Discussion of Major Unresolved Issues

It would be possible to revisit each of the identified categories of methods in quantitative geography to discuss unresolved issues and frontiers for future research. Rather than take such an approach, some overarching themes are discussed below associated with major unresolved issues in quantitative geography. The themes to be discussed are the following: spatial data uncertainty, abstraction and frame independence, spatial and spatio-temporal patterns, and inter-category integration.

A oft discussed issue is spatial data uncertainty. For quantitative geography this is no doubt important and significant. In particular, there may be potential for spatial data uncertainty (or error) to propagate through analysis in different ways, and this could vary by method. Goodchild (2008) reviews related issues and preliminary progress on this front, but it remains an unresolved issue as to exactly what data uncertainty does to geographic analyses.

A model, whether data, statistical or mathematical in orientation, is an abstraction of reality. As discussed previously, a model reflects intended (and unintended) relationships, and one's knowledge of relationships. If nothing else, we are now keenly aware that our models of reality have much potential to influence or bias analysis. For example, this is the crux of the modifiable areal unit problem discussed in Openshaw and Taylor (1981), and there is ample evidence in many application domains that model results are sensitive, or influenced, by spatial data scale and/or unit definition. Tobler (1979) highlights that model specification, or rather incorrect model specification, is a likely source of error, resulting in an inability to statistically confirm certain spatial relationships. This is referred to as frame dependence, or rather that a particular statistical test is dependent on the spatial frame utilized and a different frame may produce differing results. Somewhat related to this issue is the work by Murray (2005) highlighting that certain spatial optimization models are sensitive to spatial data scale and/or unit definition, and are therefore frame dependent. Finally, Florax and Rey (1995) demonstrated that the spatial lag model was sensitive to the relationships specified in the spatial weights matrix. Thus, at issue is that data could vary by scale and unit definition, the underlying model is but one abstraction of reality and others could exist, and parameters in models sometimes reflect limited knowledge. These remain unresolved issues in general terms, but research has already begun to address particular cases. For example, Wong (2002) introduces variants of spatial segregation measures that are less frame dependent. In spatial optimization modeling, Murray (2005) introduces a new model for set covering that is less frame dependent. Finally, Aldstadt and Getis (2006) develop an approach to make the specification of the spatial weights matrix more endogenous, and better justified.

An increasing need in geographic analysis with enhanced and great abundance of spatial information is better accounting for patterns and shapes. Not only to detect spatial patterns, as reviewed previously, but developing models that reflect patterns and/or help to produce particular patterns in the case of prescription. Of course the first step is being able to quantify aspects of pattern or shape. Thus, the work of Williams and Wentz (2008), as an example, is a noteworthy attempt to move into the direction of producing a pattern/shape, once its specification can be given. Another example is the developed measure of relative contiguity in Wu and Murray (2008), enabling patterns that are more or less contiguous to be distinguished. Of course, these are specific instances, so unresolved is a more general framework to evaluate and produce patterns using models.

Perhaps not so obvious in the review of quantitative geography methods is work focusing on spatio-temporal patterns. It most certainly is implied, but admitted not discussed in any detail. If we did discuss space-time approaches, it likely would have been in the GIS or ESDA sections, as work to date has been qualitative and visual. Thus, an unresolved issue is moving beyond the more descriptive geovisualization approaches for space-time analysis and establishing a mathematical framework for such a method. Miller (2005) offers a preliminary direction on this front.

Finally, the future offers much potential in quantitative geography for further cross fertilization and interaction of the various specialty areas. One example is the work of Ward et al. (2003) where high resolution spatial information, a regional optimization model and a sub-area cellular growth simulation model are linked together. This was done to evaluate population growth and spatial impacts, but regional decisions/changes are linked to local growth and development. Another example is the spatial optimization model detailed in Matisziw and Murray (2009) for maximizing coverage in facility siting. An approach is developed to solve the model using GIS based methods, taking advantage of spatial relationships and knowledge, thereby reflecting an intelligent approach to searching space. While there is ample evidence of inter-category integration already occurring, much remains possible. Further, important research questions will no doubt arise.

Conclusions

This paper has attempted to provide a summary of the many methods in quantitative geography. Six broad categories were used to detail the range of methods found in quantitative geography: geographic information systems; airborne sensing (global positioning system, photogrammetry and remote sensing); statistics and exploratory spatial data analysis; mathematics and optimization; regional analysis; and, computer science and simulation. Methods within these categories were discussed with respect to the contributions of geographers, so certainly do not necessarily reflect the broader contributions on particular topics by research across other academic disciplines. The broad collection of methods in quantitative geography are based upon the many contributions of noteworthy quantitative geographers, many of whom are not even cited in this paper. This was not intentional, but rather due to space limitations. Based on the above discussion and current trends, it is fair to say that the quantitative revolution in geography is alive and well. However, contributors (both in terms of methods and application) have branched way beyond geography, and regional science.

References

Aldstadt, J. and A. Getis (2006). “Using AMOEBA to create a spatial weights matrix and identify spatial clusters.” Geographical Analysis 38, 327-343.

Anselin, L. (1988). Spatial Econometrics: Methods and Models (Kluwer Academic Publishers: Dordrecht, The Netherlands).

Anselin, L. and S. Bao (1997). “Exploratory spatial data analysis linking SpaceStat and ArcView.” In Recent Developments in Spatial Analysis, edited by M. Fisher and A. Getis, pp. 35-59 (Springer: Berlin).

Anselin, L., I. Syabri, and Y. Kho (2006). “GeoDa: An introduction to spatial data analysis.” Geographical Analysis 38, 5–22.

Armstrong, M.P. (2000). “Geography and computational science.” Annals of the Association of American Geographers 90, 146-156.

Batty, M. and P.A. Longley (1994). Fractal cities (Academic Press: London).

Berry, B.J. L. and A.M. Baker (1968). “Geographic sampling.” In Spatial analysis, edited by B.J.L. Berry and D.F. Marble, pp. 91-100 (Prentice-Hall: Englewood Cliffs, NJ).

Berry, B.J.L. and D.F. Marble (eds) (1968). Spatial Analysis (Prentice-Hall: Englewood Cliffs, NJ).

Brunsdon, C. S. Fotheringham and M. Charlton (1998). “Geographically weighted regression - modelling spatial non-stationarity.” The Statistician 47, 431-443.

Casetti, E. (1972). “Generating models by the expansion method: application to geographic research.” Geographical Analysis 4, 81-91.

Church, R.L. and A.T. Murray (2009). Business Site Selection, Location Analysis, and GIS (Wiley: New York).

Clarke, K.C., S. Hoppen and L. Gaydos (1997). “A self-modifying cellular automaton of historical urbanization in the San Francisco Bay area.” Environment and Planning B 24, 247-261.

Clark, W.A.V. and P.L. Hosking (1986). Statistical Methods for Geographers (Wiley: New York).

Cole, J.P. and C.A.M. King (1968). Quantitative Geography: Techniques and theories in geography (Wiley: New York).

Fischer, M.M. (2006). “Neural networks: a general framework for non-linear function approximation.” Transactions in GIS 10, 521-533

Florax, R.J.G.M. and S.J. Rey (1995). “The impacts of misspecified spatial interaction in linear regression models.” In New Directions in Spatial Econometrics, edited by L. Anselin and R. Florax, pp. 111-135 (Springer-Verlag: Berlin).

Fotheringham, A.S., C. Brunsdon and M. Charlton (2000). Quantitative Geography: Perspectives on spatial data analysis (Sage: London).

Goodchild, M.F. (2008). “Statistical perspectives on geographic information science.” Geographical Analysis 40, 310-325.

Grubesic, T.H., T.C. Matisziw, A.T. Murray and D. Snediker (2008). Comparative approaches for assessing network vulnerability.” International Regional Science Review 31, 88-112.

Haggett, P., A.D. Cliff and A. Frey (1977). Locational Analysis In Human Geography (Wiley: New York).

Haggett, P. and R.J. Corley (1969). Network Analysis in Geography (St. Martin’s: New York).

Haggett, P. (2008). “The location shape of revolution: reflections on quantitative geography at Cambridge in the 1950s and 1960s.” Geographical Analysis 40, 336-352.

Haynes, K. and M. Dinc (1997). “Productivity change in manufacturing regions: a multifactor/shift-share approach.” Growth and Change 28, 201-21.

Hewings, G.J.D. (1985). Regional Input-Output Analysis (Sage Publications: Beverly Hills).

Huff, D.L. (1964). “Defining and estimating a trading area.” Journal of Marketing 28, 34-38.

Jackson, R.W. and A.T. Murray (2004). “Alternative formulations for updating input-output matrices.” Economic Systems Research 16, 135-148.

King, L.J. (1969). Statistical Analysis in Geography (Prentice Hall).

Longley, P.A., M.F. Goodchild, D.J. Maquire and D.W. Rhind (2005). Geographic Information Systems and Science, second edition (New York: Wiley).

Matisziw, T.C. and A.T. Murray (2009). “Siting a facility in continuous space to maximize coverage of continuously distributed demand.” Socio-Economic Planning Sciences 43, 131-139.

McHarg, I. (1969). Design with Nature (Natural History Press: Philadelphia).

Miller, H.J. (2005). “A measurement theory for time geography.” Geographical Analysis 37, 17-45.

Murray A.T. (2005). “Geography in coverage modeling: exploiting spatial structure to address complementary partial service of areas.” Annals of the Association of American Geographers 95, 761-772

Murray, A.T. and R.L. Church (1996). “Applying simulated annealing to location planning models.” Journal of Heuristics 2, 31-53.

Murray, A.T. and V. Estivill-Castro (1998). “Cluster discovery techniques for exploratory spatial data analysis.” International Journal of Geographical Information Science 12, 431-443.

Murray, A.T. and T. Shyy (2000). “Integrating attribute and space characteristics in choropleth display and spatial data mining.” International Journal of Geographical Information Science 14, 649-667.

Openshaw, S. and P.J. Taylor (1981). “The modifiable areal unit problem.” In Quantitative Geography: a British view, edited by N. Wrigley and R. Bennett, pp. 60-69 (Routledge and Kegan Paul: London).

Rey, S.J. (2001) “Mathematical models in human geography.” In International Encyclopedia of the Social & Behavioral Sciences, edited by N.J. Smelser and P.B. Baltes, pp. 9393-9399 (Pergamon: Oxford).

Rey, S.J. and L. Anselin (2007). “PySAL: A Python library of spatial analytical methods.” The Review of Regional Studies 37, 5-27.

Rey, S.J. and M.V. Janikas (2006). “STARS: space-time analysis of regional systems.” Geographical Analysis 38, 67-86.

Robinson, G.M. (1998). Methods and Techniques in Human Geography (John Wiley & Sons: Chichester).

Rogerson, P.A. (2006). Statistical Methods for Geography, second edition (Sage: London).

Taylor, P.J. (1977). Quantitative Methods in Geography (Houghton Mifflin Co.).

Thurston, J., T.K. Poiker and J.P. Moore (2003). Integrated Geospatial Technologies (New York: Wiley).

Tobler, W.R. (1979). “Smooth pycnophylactic interpolation for geographical regions.” Journal of the American Statistical Association 74, 519-536.

Tobler, W.R. (1989). “Frame independent spatial analysis.” In The Accuracy of Spatial Databases, edited by M. Goodchild and S. Gopal, pp. 115-122 (Taylor and Francis: New York).

Tukey, J.W. (1977). Exploratory Data Analysis (Addison-Wesley: Reading).

Ward, D.P., A.T. Murray and S.R. Phinn (2000). “A stochastically constrained cellular automata model of urban growth.” Computers, Environment and Urban Systems 24, 539-558.

Ward, D.P., A.T. Murray and S.R. Phinn (2003). “Integrating spatial optimization and cellular automata for evaluating urban change.” Annals of Regional Science 37, 131-148.

Williams, E.A. and E.A. Wentz (2008). “Pattern analysis based on type, orientation, size, and shape.” Geographical Analysis 40, 97-122.

Wilson, A.G. (1970). Entropy in urban and regional modelling (Pion Press: London).

Wong, D.W.S. (2003). “Spatial decomposition of segregation indices: a framework toward measuring segregation at multiple levels.” Geographical Analysis 35, 179-194.

Wrigley, N. and R.J. Bennett (eds.) (1981). Quantitative geography: a British view (Routledge and Kegan Paul: London).

Wu, X. and A.T. Murray (2008). "A new approach to quantifying spatial contiguity using graph theory and spatial interaction." International Journal of Geographical Information Science 22, 387-407.